Research on Display Image Optimization Based on Visual Models for Next-generation Interactive Media

Taiki Fukiage
Distinguished Researcher, NTT
Communication Science Laboratories

Abstract

High-quality, low-latency, and energy-saving video generation is foreseen to become a requirement in next-generation interactive media. We spoke to Dr. Taiki Fukiage, a distinguished researcher who is aiming to establish an image-generation technology that is both human- and environment-friendly through the construction of visual models.

Keywords: visual model, vision science, media technology

What is "display image optimization based on visual models?"

—What kind of research are you doing?

My research would probably be best described as "media technology research based on vision science." I am engaged in research to improve media display technology and come up with entirely new display methods by taking advantage of the human visual characteristics elucidated by vision science. In this context, vision science is, roughly speaking, a field of study that elucidates how information seen through the eyes is processed in the brain and how it generates subjective perceptual experiences.

Applying vision science to media technology often necessitates predicting the perceptual experience that can be gained from any image or video. However, the main approach for research in the field of vision sci-

ence thus far has been to break down stimuli into elements such as color, shape, and motion, and examine how each of them are processed. Thus, there are still no models to generically explain the way we see natural images and videos, as we see them in daily life. Therefore, the need to address the unsolved issues of vision science usually arises in considering technological applications.

To develop the relevant media technologies, we are conducting research by first figuring out the unsolved parts of vision science through experiments, and then building visual models to optimize the display images. To look into "how human visual information processing is carried out," we are working on both the scientific aspect of vision science for building models for visual systems and the technical aspect for optimizing the models by applying them to media display technology, enabling human- and environment-friendly video-generation technology (**Fig. 1**).
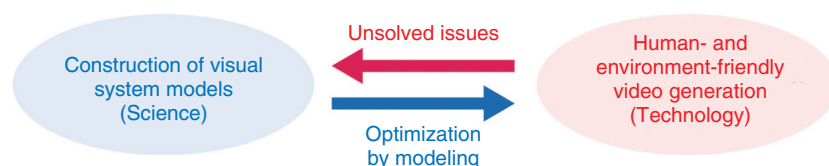
Fig. 1.   Display image optimization based on visual models.

—Specifically, what media technologies are you researching?

I would like to tell you specifically about three media technologies.

HenGenTou (2015–2019)

Research on HenGenTou (deformation lamps) started before I joined the Laboratories and pertains to a technique that uses projection mapping to add movements to stationary objects. Previous research in vision science has shown that humans recognize the world around them by independently processing motion, shape, and color information, and subsequently integrating these different kinds of information. The research started from the idea that we could do interesting things if we took the process backwards. It is a technique that uses human illusion to make stationary objects appear to be moving; and I was very surprised myself when I saw it for the first time.

Creating the illusion that an object appears to be moving requires that the projection pattern be aligned exactly with the actual object. I was mainly responsible for the component pertaining to automatic calibration of position using computer vision technologies. Also, because it is an illusion-based technique, there is a limit to the amount of motion that can be added. If you try to make very large movements, the illusion would be seen as "a clearly different pattern is added on a stationary object." Therefore, we built a "projection dissonance model" that predicts "how much movement can be permitted" and worked on research that automatically gives the maximum movement within the permitted range.

Having visual impact and appeal, the technique has been commercialized in a variety of fields. For example, it has been used to add movements to point-of-purchase advertising at stores and applied to art works.

Hidden Stereo (2016–2017)

Hidden Stereo is a technology related to three-dimensional (3D) television. On a 3D television, where images appear in 3D with the use of special glasses, you can only see blurry images without the glasses on. Using the mechanism by which humans perceive depth, we developed a stereoscopic image-generation technology that enables users to see images clearly even without 3D glasses (**Fig. 2**).

To perceive depth, humans use a mechanism to resolve the images of the left and right retinas into elements such as orientation and fineness/roughness, and then detect the phase difference between the corresponding elements as disparity in order to perceive depth. The minimum necessary induction pattern for humans to feel depth is then created on the basis of this disparity detection mechanism, to create the image for the left eye by adding that pattern to the original image, and for the right eye by subtracting it. With 3D glasses, images appear three-dimensionally because left-eye and right-eye images are delivered respectively to the left eye and the right eye. But, without 3D glasses, the plus pattern for the left eye and the minus pattern for the right eye cancel each other out, leaving only the original image, so that one can see the image clearly.

This research was originally started from the idea that depth could be added to HenGenTou mentioned above. Their similarity is that HenGenTou add patterns to give movement, while Hidden Stereo adds patterns to give depth.

Intuitive image blending based on visibility (2020–2021)

"Intuitive image blending based on visibility" is our most recent research. There is demand for technologies for blending and displaying translucent images, such as for superimposing translucent images to render emotional scenes and stacking translucent information to prevent real-world scenery from being hidden in augmented reality (AR), virtual reality (VR), etc.

Left image $L = I + D$

Disparity inducer pattern D

Original image (Middle viewpoint) I

Disparity inducer pattern $-D$

Right image $R = I - D$

L

R

$2I$

3D scene

2D scene

Image quality when viewed without 3D glasses

Conventional stereoscopic images

Proposed method

Overlapping the left and right images enables viewers without 3D glasses to perceive the original 2D image where the disparity inducer patterns have been canceled out.
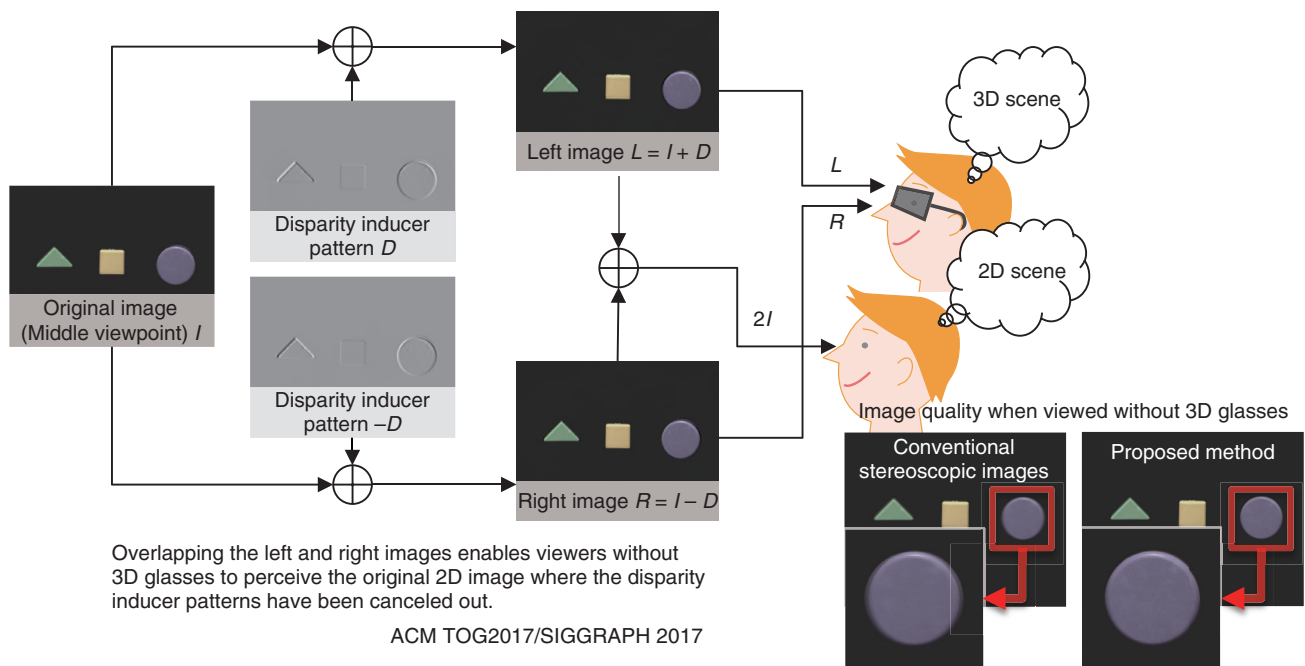
ACM TOG2017/SIGGRAPH 2017

Fig. 2.   Hidden Stereo.

This type of image blending requires setting the transparency level. However, even if the transparency is set uniformly to 0.5, visibility would completely differ depending on the combined images. It is difficult to get the desired results by simply synthesizing images by physically setting the transparency level.

Although models have been created to predict the visibility of individual elements, such as the color and coarseness of images, it has not been clear how to integrate these elements in order to predict the actual visibility of natural images. In this research, we are developing a "visibility prediction model" that can accurately predict visibility by measuring the visibility of blended images using a large number of natural images, and then automatically setting the optimum transparency level.

Aiming to establish video-generation technologies as well as contribute to vision science

—*What are the directions of your future research?*

Thus far, we have focused on applying the low-order visual science models, i.e., those near the entry point of the visual system, to media engineering. There are still many possible areas to explore in this field, but I feel that we have been able to clearly grasp

what kinds of initiatives are doable.

Therefore, going forward, we will try to model the processing of higher-order information and optimize the models more aggressively. For example, previous studies have basically predicted the way we look through the central vision, near the so-called "central fovea," but we would like to also work on predicting the way we look through the "peripheral vision." In fact, a large part of the human vision involves peripheral vision. Furthermore, how information is processed for peripheral vision is considerably different from how it is processed for central vision. I believe that modeling this properly would lead to many new possibilities.

—*What is your ultimate goal?*

The three technologies presented here differ in the details, but their final goal is the same. There are two focal goals; the first is contributing to vision science. With a view to technological applications and a wide range of images and videos treated as stimuli, these studies may lead to the discovery of important visual functions that we have missed in the past.

And the second is contributing to technology. We aim to solve technical problems in the field of media engineering. Compared with previous displays,

next-generation interactive media, such as AR and VR, require considerably more information and better quality, so it will likely become important to carefully distribute costs in consideration of human characteristics. On the basis of the constructed visual model, if we can determine the essential information necessary for eliciting natural human feelings, and then trim down unnecessary information, we will be able to achieve high-quality, low-latency, and energy-saving "human- and environment-friendly" video generation.

—Could you give a message to future business partners and young researchers?

NTT Communication Science Laboratories is a research laboratory that conducts basic research on an extremely wide range of fields. From only what I know, various studies are being conducted from brain science to computer vision, speech recognition, signal processing, and theoretical mathematics, the foundation of information processing. For many years, the fact that we have been conducting basic research without focusing on business applications is unusual for corporate research laboratories; but I think this is a major strength of NTT. In addition, while the degree of freedom for research is high, like that in universities and research institutes, we are not subject to many budget constraints. I would like to make use of this environment and pursue research with a broad perspective and free thinking, while creating technology seeds that will help society in the future.

If you want to become a researcher and are interested in the research topics mentioned above, I invite you to join our team and work with us. NTT hopes to provide an environment where you can maximize your full potential if you are engaged in your research and able to obtain competitive skills and knowledge while at university. It may also be important to find the topic that stimulates your curiosity the most. This is because a large part of research involves endless steady tasks that often do not go as planned; and I think curiosity will serve as the driving force for research during these seemingly fruitless times. As a researcher, I know that I still have a long way to go, and oftentimes I lose confidence. But ever since I was a student, I have had a strong interest in how humans perceive the world visually, and my curiosity about this problem has been the driving force for me to persevere in research.

■ **Interviewee profile**
Taiki Fukiage joined NTT in 2015, and is currently a member of the Sensory Representation Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories. He has been a distinguished researcher at the Laboratories since 2020. He is conducting research on optimizing display images using visual models by leveraging expertise in the two fields of vision science and media engineering.